基于多层卷积神经网络的金融事件联合抽取方法*

■ 李旭晖[1,2]　程威[1]　唐小雅[3]　于滔[1]　陈壮[3]　钱铁云[3]

[1] 武汉大学信息管理学院　武汉 430072　[2] 武汉大学大数据研究院　武汉 430072

[3] 武汉大学计算机学院　武汉 430072

摘　要：[目的/意义]为进一步提升金融领域事件抽取的效果，增强事件抽取两个子任务之间的关联性。[方法/过程]在中文金融文本上进行事件抽取相关研究，提出一种融合预训练模型与多层卷积神经网络的金融事件联合抽取方法，首先通过预训练模型 BERT 捕捉句子序列的综合语义信息，然后接入本文设计的多层卷积架构 MultiCNN，分层提取局部窗口和高维空间语义信息，同时实现事件识别和要素抽取这两个任务，再通过引入对比损失，进一步强化两个任务之间的关联。[结果/结论]在中文金融事件数据集上 F_1 达到 82.20%，比各个基准抽取模型均有一定提升。

关键词：中文事件抽取　卷积神经网络　预训练模型　联合学习

分类号：G250

DOI：10.13266/j.issn.0252－3116.2021.24.010

1 引言

事件是人们表达信息、观点或事实的一种语义单元呈现，它一般是指特定的人或物在特定时间地点内相互作用的客观事实，是人们认知世界的一种方式[1]。金融行业有以新闻、公告、微博等为代表的海量领域文本数据，这些文本中蕴含着以金融事件为代表的高时效、高价值信息，如投融资事件、收购重组事件等。这些事件能用于辅助企业进行信息决策，同时还能为构建金融事件体系及知识图谱提供底层技术支持，具有重要应用价值。如何有效获取这些事件信息一直是业界与学界共同关注的问题，其主要途径是对领域文本进行事件抽取。

事件抽取是从自然语言文本中识别特定的事件类型，同时抽取出与事件表达相关的时间、地点、人物等要素，最终以结构化的方式呈现出来[2]。通常来说，事件抽取包括两个子任务：①事件检测与分类；②事件要素识别与分类。事件抽取是自然语言处理中的热点研究问题，近年来得到了学术界的广泛关注，在信息抽取方向的主要评测会议 ACE（automatic content extraction）中专门设定了事件抽取的评测分支。ACE 定义了共计 33 种事件类型，涉及出生死亡、公司成立、交通肇事、入狱释放等多样化领域，并提供了多种语言的平行标注语料供评测[3]，目前大多数研究均针对 ACE2005 的语料进行优化与实验。

金融事件抽取旨在通过对金融文本进行快速且准确的信息抽取，根据业务需求提取出其中关键的事件信息。金融事件抽取有着较为明显的领域需求特点，主要体现在事件类型相对专业化、事件要素识别准确率要求较高这两个方面，直接移植传统事件抽取会存在以下几点问题：①金融事件类别分布稀疏，如 ACE2005 仅包含 633 篇中文文档，涵盖 33 类事件，训练语料过少且很多类别与金融领域关注的事件无关（如死亡、攻击、示威游行等），并且事件类别过多会导致抽取模型存在较为严重的过拟合；②事件句分类和事件要素抽取准确率不足，抽取模型需要进一步优化；③事件分类与事件元素抽取两个子任务独立建模，任务之间易发生错误传播。因此金融领域的事件抽取任

* 本文系国家自然科学基金重大研究计划“大数据驱动的管理与决策研究”重点支持项目“基于知识关联的金融大数据价值分析、发现及协同创造机制”（项目编号：91646206）和深证信息联合研究计划课题“企业全生命周期关键事件识别和要素抽取”（项目编号：CHINFO201802）研究成果之一。

作者简介：李旭晖，副教授，硕士生导师，E－mail：lixuhui@whu.edu.cn；程威，硕士研究生；唐小雅，硕士研究生；于滔，硕士研究生；陈壮，博士研究生；钱铁云，教授，博士生导师。

收稿日期：2021－06－15　修回日期：2021－09－27　本文起止页码：89－99　本文责任编辑：易飞

务需要有更具针对性的语料积累和抽取方法。

基于上述原因,本文聚焦于金融领域事件抽取的实际需求,深入探讨了其中的关键技术问题,并提出了新的金融事件抽取方法。本文主要工作集中体现在以下三个方面:①建立金融领域事件分类体系,收集金融资讯网站的真实新闻语料,经预处理后按照事件抽取任务的需求进行字级别的 BIO 标注;②针对金融领域文本的特点,提出了一种融合预训练语言模型 BERT[4]与多层卷积神经网络[5]的事件联合抽取方法 BERT-MultiCNN,分层捕捉句中的事件语义信息,进一步提升事件抽取效果;③引入两阶段联合学习机制来处理多任务协同问题,解决流水线模型中的错误传播问题,并通过对比损失建立了任务间的合理关联,使得第一步事件分类的错误可以通过共享下一步要素抽取任务中的信息来进行纠正。论文设计了相关对照实验及探索实验,实验结果表明,本文提出的模型在中文金融事件数据集上 F_1 可达到 82.20%,能够进一步提升现有金融事件抽取两阶段子任务的效果,具有较高的应用价值。

2 相关研究

事件抽取是自然语言处理中的关键任务之一,在舆情分析、语义理解、文本摘要等方面应用广阔,国内外诸多学者在此方面均进行了广泛的研究,经整理,事件抽取目前的研究可以大致分为三大类:基于模式匹配的方法、基于特征构造的机器学习方法、基于深度学习的方法。

模式匹配的方法,通常是在领域专家的帮助下设计规范的抽取规则或者模板,其核心在于模式的构建。如 E. Riloff 等提出的 AutoSlog[6],以及后面的改进版 AutoSlog-TS[7],可以利用模式词典构建模板从而实现自动抽取,还能根据现有模式去学习新的模式,在一定程度上降低了人工构建规则的成本。在金融领域,R. Feldman 等[8]通过构建金融情感词典,实现短语级别的情感极性匹配并制定了事件抽取规则;罗明等[9]通过定义金融领域的事件表示模型,使用词向量工具自动生成同义词,然后基于规则模式实现多种金融事件的抽取。总结而言,利用模式匹配的方法通常需要大量规范的抽取规则,构建成本较高且不具有通用性,因此在应用时存在较大局限性。

随着机器学习技术在语音、翻译等领域的广泛成功,诸多学者逐渐将目光转向了利用机器学习的方式进行抽取。这种方式通过构建词法、句法、词位置等多种特征,将其作为机器学习模型的输入从而实现自动化的事件抽取,常见的机器学习模型有支持向量机(SVM)、条件随机场(CRF)、隐马尔可夫模型(HMM)等。李响与杨小琳等[10]通过设计词法、句法和语义三类特征,然后利用支持向量机实现了新闻事件的分类,L. Hou 等[11]利用条件随机场进行机器学习模型的构建,通过融合词汇特征、语义、依存关系、句法关系、相对位置这 5 类特征,并加入语义角色特征实现事件抽取。基于机器学习的方法难点在于特征的选择以及组合,不同学者之间研究的差异也主要体现在特征的选择以及机器学习模型的选择方面,因此这种方式难以避免地会出现主观偏差。

与传统机器学习的方式不同,利用深度学习做事件抽取的任务,一般逻辑是将文本序列表示为一段可计算的多维张量,通过构建端到端的深度学习模型实现事件触发词与事件要素的分类。事件句分类则等价于对该张量进行 k 分类(k 为事件句的类别 +1,多出来的一类表示非事件句),若要实现事件元素的抽取,则需要对每个位置的向量表示进行类别总数为 t 的多分类(t 为所有事件元素类别数 +1,多出来的一类表示非事件元素标签)。这种方式通常不需要事先进行特征工程,可以降低主观意愿对模型输入的影响,实际使用中,卷积神经网络(CNN[5,12-15])、循环神经网络(RNN[13-17])、Transformer 编码器[18-22]以及图神经网络[19,23-24]是应用最为广泛的特征抽取器。如 Y. Chen 等[12]最先利用深度神经网络架构进行事件抽取任务,通过设计动态多池化卷积神经网络(DMCNN),融合了词向量、位置向量以及实体类型向量,完成事件分类与要素抽取两大任务;T. H. Nguyen 等[16]利用 Bi-GRU 对句子进行编码,通过设计不同的记忆矩阵对事件分类、要素识别这两大任务进行联合抽取,减少了错误传播;S. Zheng 等[19]为了解决事件论元分散在多个句子中的问题,提出了一种端到端的解决方案 Doc2EDAG,通过生成基于实体的有向无环图,使用多个 Transformer 对图进行编码;图模型层面,不少学者通过将句子序列视为一个有向无环图,利用图神经网络模型实现事件的分类抽取,如 T. H. Nguyen 等[23]提出一个基于句法依存树的图卷积(GCN)神经网络用于事件检测,后来 S. Cui 等[24]在其基础上,提出了一种新型的关系感知 GCN (RA-GCN)结构,研究了不同句法关系标签对事件抽取的影响,进一步优化了前面的工作。横向对比来看,几大基于深度学习的方法实现事件抽取,其优缺点对比如表 1 所示:

李旭晖，程威，唐小雅，等.基于多层卷积神经网络的金融事件联合抽取方法[J].图书情报工作,2021,65(24):89－99.

表1　基于深度学习实现事件抽取的方法对比

编码器	优点	缺点
CNN	(1)参数共享机制利于模型训练; (2)可以捕捉局部窗口信息; (3)池化方式可多样化,以优化句向量表达	(1)网络结构浅层导致语义编码能力有限; (2)缺少位置编码信息,对于词级别分类有较大损失; (3)卷积结构单一,对于事件元素抽取类任务准确率不高
RNN	(1)序列方式建模符合对序列文本的直观理解; (2)多种 RNN 变种可以存储上下文信息	(1)仍然存在短期记忆问题,对句向量表示有待优化; (2)每个时间步逐步输出,计算效率低
Transformer	(1)语义编码能力更加强大、高效; (2)基于注意力机制,可以建模任意两个位置的关联信息; (3)基于预训练架构,无监督语义信息更加丰富	(1)参数量太大,在一定程度上影响模型效率; (2)预训练模型无法解决一词多义问题
Graph	(1)非欧空间建模,有助于捕捉语句中的非连续依赖关系; (2)天然适合建模句法层面的事件元素依赖关系	(1)参数量大,训练慢且易发生过平滑; (2)需要进行分词及句法分析,前后任务存在较大错误传播问题
混合编码器	(1)可以分别利用各个编码器的优势,取长补短; (2)可以对两个任务进行分层建模,分步优化	(1)存在错误传播问题; (2)存在多任务协同问题

将机器学习和深度学习应用到金融领域的事件抽取任务,也一直是学术界和工业界重点关注的问题。除了前文所述的模型层面的探索与使用,金融事件抽取还包含以下两个应用方面的研究:①语料获取方面,有监督的机器学习方法需要较为成熟的标注语料,H. Yang 等[25]通过远程监督的方式获取更多的金融领域标注语料,然后利用 BiLSTM-CRF 实现句子级及篇章级的事件抽取;L. Ein-Dor 等[26]则利用维基百科数据,自动化提取金融事件中的弱标签,并利用相关实验证明了该方法的有效性;Z. Zhou 等[27]提出了一个高低双层事件检测模型,用于识别篇章级企业金融事件,同时基于新闻语料开发了一个新的数据集 EDT,用于辅助企业事件识别与股票价格预测任务;②输出结果方面,金融事件抽取一般需要有明确的应用目标,S. Rönnqvist 等[28]利用深度学习模型抽取新闻中的金融风险事件,并利用抽取结果对银行危机指数及政府干预情况输出分析报告;S. Carta 等[29]利用层次化聚类的方法对金融事件进行分类,一方面向用户提供每日新闻的结构化摘要信息,另一方面可以提醒用户一些可能发生的潜在重要事件;L. D. Corro 等[30]通过设计一个独特的注意力网络处理大量新闻标题数据,用来预测给定股票的价格变动情况(涨/跌/平),可以在不需要标记数据的条件下达到较好的股票预测效果。

综上可知,早期事件抽取的研究主要集中在基于规则构建模板的方式,这种方式需要大量的人工规则,且不具有通用性。随着深度学习模型结构的发展,最新研究主要通过深度学习进行事件抽取,由于预训练模型具有丰富的语义表达能力,在预训练模型基础上搭建个性化的抽取模型,并通过联合训练的方式微调模型参数也逐渐成为完成事件抽取的主流方法。然而目前基于预训练模型的事件抽取相关研究大都将语义捕捉的任务交予普适性较强的预训练模型完成,对单个句子局部信息的捕捉缺乏针对性。卷积神经网络(CNN)利用卷积核共享机制,可以充分利用局部窗口内的上下文信息对特征进行提取,在文本分类[5,31]、序列标注[32-34]、事件抽取[12-14,35]等方面均取得了很好的效果,然而将这些研究移植到事件抽取这一任务中时,可能存在以下几点问题:

- 特征提取方面:以往研究大都通过叠加词向量、词法、句法等维度特征,并用 CNN 进行特征提取与整合,而受限于 CNN 网络结构本身的浅层,其语义编码能力有限;
- CNN 结构方面:卷积结构过于单一,只能通过优化卷积层数、改进池化方式等方式优化整体架构;
- 利用 CNN 分别卷积,独立进行事件分类和事件要素抽取,会造成两个任务之间的错误传播。

基于此,本文结合预训练模型与 CNN 两者的优点,先利用预训练语言模型 BERT 进行通用语义信息提取,再设计独特的多层卷积神经网络用于分层捕捉局部信息,并对两阶段任务进行联合端到端训练,最终实现事件句分类和要素抽取任务。

3　金融事件联合抽取方法介绍

本文提出一种金融事件联合抽取方法,该方法可有效解决事件抽取中错误传播的问题。为了获取金融领域的典型事件,先构建了金融事件分类体系;为了进行模型优化与抽取方案有效性的检测,选择了收购重组、战略合作、投融资这3类重点事件进行语料标注,并在此基础上进行相关优化及探索实验。

3.1 金融事件分类体系

结合对金融领域的背景调研，本文初步建立了金融事件分类体系，并从公司金融的角度对事件进行分类，完整分类后的事件类别及其细分事件名称如表 2 所示。同时，为了针对具体事件类别获取广泛的抽取语料，我们利用关键词抽取技术对事件特征词进行提取，并利用词向量工具 Word2Vec[36] 进行特征词的扩展，根据扩展后的特征词集合搜集大量的中文事件抽取语料。

表 2 金融事件分类

事件类别	事件名称
并购重组	公司收购、资产重组、混改、撤销重组
财务	业绩亏损、中标、净利、购买银行理财产品、注销专项账户、增值税退税
公司治理	资金占用、离职辞职、减持股份、撤销职务、召开董事会、权益分派、召开临时股东大会、变更实际控制人
股票	上涨、下跌、停牌、复牌、涨停、跌停、股价异动、股份转让、股权质押、定增、非公开发行股票、发行限售股份、股权解押
监管	关注函、违规举牌、回应函、行政处罚、信息披露违法、列为失信被执行人、问询函、处以罚款、证监会点名、反垄断罚款、列入黑名单、监管函、行政监管、强制退市、强制摘牌
债务	债务违约、企业负债、账户冻结、账户解冻、强行划款、借款、银企合作
违法违规	非法集资、股东虚假出资、票据违法、违规担保、虚报注册资本、隐瞒关联关系、操纵市场
其他	股票简称变更、撤销登记、工商变更、债券名称变更、协议违约、战略合作

3.2 问题定义

金融领域的事件抽取工作的任务一般定义为：从非结构化的文本中自动化地抽取出关键事件，以及该事件包含的事件元素（如时间、地点、事件主体等），得到关于事件的结构化表述。如下面例句所示，模型需要将其判定为战略合作事件，并抽取出事件触发词（战略合作）、合作公司（海天水务集团）、合作公司（中信银行成都分行）、事件发生时间（4 月 6 日）、事件发生地点（成都）这些事件要素，具体到每个字符的标签空间，可参照表 3 所示。

例句：4 月 6 日，海天水务集团与中信银行成都分行在成都签署战略合作协议。

表 3 事件抽取字符与正确标签对应关系

4	月	6	日	，	海	天	水	务	集	团
B-time	I-time	I-time	I-time	O	B- cpn	I- cpn	I- cpn	I- cpn	I- cpn	I- cpn
与	中	信	银	行	成	都	分	行	在	成
O	B- cpn	I- cpn	I- cpn	I- cpn	I- cpn	I- cpn	I- cpn	I- cpn	O	B-loc
都	签	署	战	略	合	作	协	议	。	
I-loc	O	O	B-trig	I-trig	I-trig	I-trig	O	O	O	

针对事件句分类与事件元素识别这两个阶段的子任务，目前存在两种主要的处理方式，即串行（Pipeline）和联合（Joint）事件抽取，串行的方式首先进行事件类型的检测，然后从事件句中抽取出相应的事件要素，而联合抽取则是同时完成事件识别与要素抽取这两个任务。由于串行方式的架构易发生误差传播，即第一阶段事件句分类的错误会影响到下一步的事件元素识别，本文采用联合抽取的方式对事件抽取的两个任务进行建模。具体而言，将事件句识别的任务视为一个句子分类任务，事件触发词以及事件元素的识别视为一个字符级别的序列标注任务，通过构建端到端的深度学习模型对事件抽取任务进行建模。

3.3 方法概述

为了完成金融事件抽取中事件分类与事件元素抽取这两个子任务，本文提出了一种基于多层卷积神经网络的金融事件联合抽取方法 BERT-MultiCNN。首先利用 BERT 强大的语义编码能力对句子序列进行向量编码，得到每个 token 位置的隐层向量编码 H_1 以及句子向量编码 C_1，然后将 H_1 通过多层卷积神经网络架构，进一步提取句子的局部语义信息，得到向量编码 H_2，然后将 H_2 输入全连接层的分类模型，抽取得到每个位置的序列标签分布，最后利用 CRF 层控制标签输出，得到事件要素和触发词的抽取结果。为了达到联合抽取的效果，一方面将其设置为端到端的一个统一架构，使两个任务损失优化的方向得到统一；另一方面，对后一个任务做序列标注的同时，对 H_2 做 Pooling 操作得到句子向量编码 C_2，与前面的句子向量编码 C_1 融合得到事件分类的结果，并计算对比损失，该损失共同形成优化目标。模型整体架构如图 1 所示，由输入层、BERT 编码层、MultiCNN 层、全连接层、CRF 层以及前后事件句表示的对比损失所组成。

本文模型创新性主要体现在以下几个方面：①融合预训练模型和传统卷积架构的共同优点，分别应用于事件抽取的两个任务，使其分别作用于句子及 token 级别的优化；②引入联合抽取模式完成事件抽取两步骤，解决错误传播问题；③引入事件句的句向量对比方案，以进一步优化多任务协同问题。

3.3.1 BERT 编码层

本文利用谷歌发布的预训练模型 BERT[4] 作为词向量编码器，BERT 模型的主要构成单元为 Transformer[18] 的 Encoder 部分，通过堆叠多层双向的 Transformer 结构，可以用来捕捉句子序列中的双向深层次语义信息，BERT 输入向量包含三个维度：Token Embeddinng

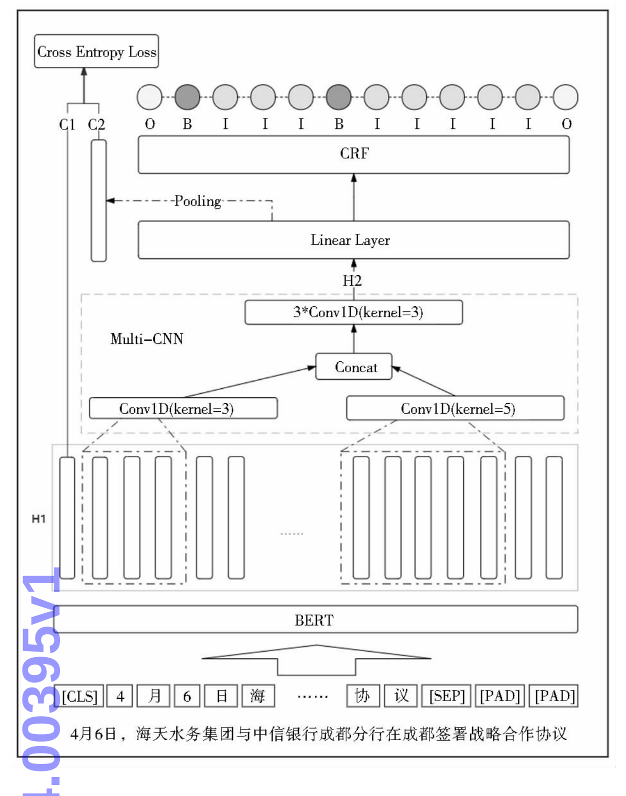


图 1　模型整体架构

表示词向量，可以使用 one-hot 编码或者使用常见的公开词向量（如 Word2Vec[36]、Glove[34] 等），对于中文来说一般是针对单字的字向量；Position Embedding 表示位置向量，采用相对位置编码对序列中每个字符的位置进行建模；Segment Embedding 表示句间分隔向量，用来标识输入的部分属于哪一个句子。BERT 的输入向量由这三部分直接相加得到：

$$E_{input}=E_{token}+E_{position}+E_{segment} \quad 公式(1)$$

得到 BERT 模型输入后，进入 Transformer 结构的 Encoder 部分，该部分主要由多头自注意力机制（Multi-head Self-attention）以及前馈神经网络组成。自注意力机制可以让模型学习到序列中两两位置之间的关系，并赋值不同的权重，"多头"意味着同时利用多个视角进行计算，以便关注句子序列中不同角度的信息，最后将来自所有注意力头的信息进行拼接，计算细节如公式(2)－公式(4)所示：

$$Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad 公式(2)$$

$$head_i=Attention(QW_i^Q,KW_i^K,VW_i^V) \quad 公式(3)$$

$$MultiHead(Q,K,V)=Concat(head_1,head_2,\cdots,head_h)W^O \quad 公式(4)$$

这里的前馈神经网络为全连接层，被单独应用到序列的每一个位置，具体而言，包括两个线性转换函数，中间使用 Relu 函数加以激活。

$$FFN(x)=max(0,xW_1+b_1)W_2+b_2 \quad 公式(5)$$

另外，Transformer 结构中的多头自注意力机制以及前馈神经网络之后都加入了残差连接和层归一化（Add & Norm）的操作，可以使得训练过程更加容易，同时能很好地避免梯度消失问题。经过 BERT 编码后，取最后一层的输出作为最终的编码表示，其中句子序列的每一个位置都对应一个维度为 768 维的隐层向量，最前面的[CLS]位即表示句子分类向量，接一个全连接层后用于事件句分类，用 C_1 进行表示。

$$H_1=BERT(E_{input}) \quad 公式(6)$$

$$C_1=FC_1(H_1[0,:,:]) \quad 公式(7)$$

3.3.2　MultiCNN 层

卷积神经网络（CNN）是一种特殊的深度前馈神经网络，最初被广泛应用于计算机视觉领域，其思想来源于人类的视觉系统，借鉴其中的感受野（receptive field）概念，从而可以接受来自不同区域的信号，对其进行整合然后作为自己的特征图输出[37－38]。后来被相关领域的学者逐渐应用到文本处理领域，如文本分类、序列标注等[5, 12－13, 32]，同样也取得了非常不错的效果。在实际使用中，一般由若干卷积层、池化层、激活层组合而成。

在事件抽取的触发词识别、事件要素识别任务上，本文将其视为一个字级别的序列标注任务，用 B、I、O 标签分别表示触发词和事件要素的实体起始位置（Begin）、实体内部（Inside）和非实体部分（Outside），这样可以对句子的所有位置同时进行分类，提升训练效率。本文利用一维卷积核在句子序列的方向上进行滑动，捕捉每个位置字的语义信息及其上下文窗口特征，最后综合得到的特征图（feature map）作为每个位置的向量输出，不同大小的卷积核可以捕捉的局部窗口大小不同。为了保持序列长度的不变性，我们对不同大小的卷积核设计了对应的 padding 操作，用于填充句子的首尾部分，长度为 3 的卷积核提取特征的过程见图 2。

具体而言，本文进行了局部窗口卷积、语义信息提升两个阶段：①局部窗口卷积上，为句子序列分别设计了 kernel 为 3 和 5 的两种卷积操作去并行化处理，然后将两者得到的特征图进行堆叠；②语义信息提升阶段，接上三层卷积核大小为 5 的卷积操作，用来提取高层语义空间信息，整体架构如图 1 所示。最后，卷积层的输出结果进入全连接层进行维度的转换，得到最终标签分类概率 P_{LABEL}。

$$H_2=MultiCNN(H_1) \quad 公式(8)$$

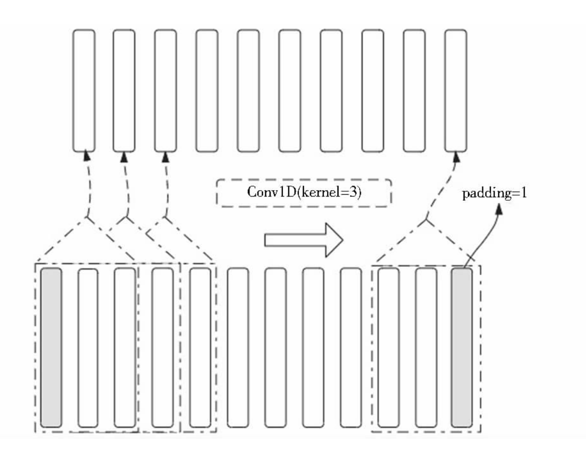


图 2　序列卷积

$$P_{LABEL} = FC(H_2) \quad \text{公式(9)}$$

3.3.3　CRF 层

事件触发词识别、事件要素识别均可以视为一个标签分类任务，而传统的 softmax 分类器未考虑到标签之间的依赖与转移概率，因此容易出现标签偏差问题。条件随机场(CRF)可以对序列标签之间的依赖关系进行建模，在词性标注、分词、命名实体识别等任务中取得了很好的效果[39-40]。CRF 可以对整个句子的标签序列进行联合建模，我们利用 CRF 层对标签输出进行控制。假设输入的句子序列为 $X = \{x_1, x_2, \cdots, x_m\}$，预测输入的标签序列为 $y = \{y_1, y_2, \cdots, y_m\}$，则整个句子序列的综合得分由节点得分(标签分数)和路径得分(标签转移分数)相加得到：

$$S(X, y) = \sum_{i=0}^{m} T_{y_i, y_{i+1}} + \sum_{i=1}^{m} P_{i, y_i} \quad \text{公式(10)}$$

其中 P 表示节点得分，即将第 i 个位置的输入预测为 y_i 标签的概率，可以由上游输出的发射概率矩阵得到；T 表示路径得分，即由 y_i 标签转移到 y_{i+1} 标签的概率，如果标签的数量为 n，则 T 为一个 $n * n$ 的矩阵，该矩阵需要在训练过程中进行拟合。我们的训练目标是正确路径的得分最大化，对所有路径得分做全局归一化后，得到正确路径的归一化概率：

$$P(y^* \mid X) = \frac{e^{s}(X, y^*)}{\sum_{\tilde{y} \sim \in Y_X} e^{s(X, \tilde{y})}} \quad \text{公式(11)}$$

训练过程中的优化目标便为正确路径的最大似然，即 $\log P(y^* \mid X)$，预测时，根据句子标签序列的预测概率进行维特比解码，即可得到最终的正确标签。

3.3.4　训练损失

在事件句分类子任务中，我们使用了交叉熵损失(Cross Entropy Loss)，用来衡量正确分类与预测分类之间概率分布的差距，假设正确的事件类别分布为 Q_1，预测的事件类别分布为 $P_{sentence}$，则事件句的分类损失 $Loss_1$ 定义为两个概率分布的交叉熵：

$$P_{sentence} = softmax(C_1) \quad \text{公式(12)}$$

$$Loss_1 = H(P_{sentence}, Q_1) = -\sum Q_1(x) \log_2 P_{sentence}(x) \quad \text{公式(13)}$$

事件要素识别的分类损失 $Loss_2$，为了将联合学习模型的损失进行同步训练，定义为 CRF 层最大似然的相反数：

$$Loss_2 = -\log P(y^* \mid X) \quad \text{公式(14)}$$

在对事件分类、事件要素抽取这两个子任务进行联合学习的过程中，需要对这两个任务进行关联，提高模型的泛化能力。一方面，对上述两个任务的损失进行相加从而达到信息共享的效果；另一方面，为了进一步融合两个子任务的语义信息，我们为卷积前后的句向量概率分布增加了一个对比损失 $Loss_3$。具体而言，本文对卷积层输出的隐层向量 H_2 进行均值池化后，得到句向量概率分布 C_2，对比损失定义为句向量概率分布 C_1 和 C_2 的交叉熵，直观上理解，通过缩小卷积前后的句向量表示差距，可以使得模型对于两个子任务的学习朝着同一个方向更新，从而强化多任务学习的交互。

$$Loss_3 = -\sum C_2(x) \ log_2 C_1(x) \quad \text{公式(15)}$$

联合学习的训练损失定义为三种损失的加权和，权值参数的比例根据实验调整得到：

$$Loss = Loss_1 + Loss_2 + 0.5 * Loss_3 \quad \text{公式(16)}$$

4　实验与分析

4.1　数据来源

本文用于金融领域事件抽取的语料为自建语料，利用 Python 爬虫从新闻资讯网站财新网(https://www.caixin.com/)、每经网(http://www.nbd.com.cn/)上爬取原始篇章级数据，首先以句子为语料单元进行切分、去重、去停用词操作，为了得到一批成熟的高质量事件标注语料，基于开源标注系统 brat(http://brat.nlplab.org/)，组织开展事件句以及事件触发词、事件要素的标注工作，整体标注流程采用 BIO 字级别标注模式，其中 B 表示 begin，即一个实体(触发词或事件要素)的开始，I 表示 inside，即一个实体的非起始位置，O 表示 outside，即无关字符，如对于句子序列“4 月 6 日 海 天 水 务 集 团 与 中 信 银 行 成 都 分 行 在 成 都 签 署 战 略 合 作 协 议”，进行标注后得到“1 B-time I-time I-time I-time B-cpn I-cpn I-cpn I- cpn I-cpn I-cpn O B-cpn I-cpn I-cpn I-cpn I-cpn I-cpn I-cpn I-cpn O B-loc I-loc O O B-trig I-trig I-trig I-trig O O”，最前面标

李旭晖，程威，唐小雅，等.基于多层卷积神经网络的金融事件联合抽取方法[J].图书情报工作,2021,65(24):89－99.

注的1表示事件句类型(用0表示非事件句)，后面每个位置对应一个字级别token的标签。另外，为了尽可能减少标注错误，我们前期组织了标注工作及背景知识的统一培训，标注上采取了2人平行标注、1人标注审核的方式。最终标注完成的语料统计后如表4所示，按照8∶1∶1的比例分别构建训练集、验证集和测试集数据。

表4　事件语料统计

事件名称	标记语料/篇	句子数量/句	事件句数量/句
战略合作	720	16 074	1 112
收购重组	650	20 058	2 228
投融资	156	14 251	1 400
总计	1 526	50 383	4 740

4.2　评价标准与环境配置

本文采用准确率(P)、召回率(R)和这两者的调和平均值(F1)对模型进行评价。当事件句分类正确，且对应事件元素抽取也正确时，判定为模型对事件要素识别正确。实验环境配置如表5所示：

表5　实验环境配置

实验环境	环境配置
操作系统	Ubuntu16.04.7 LTS x86_64
CPU	AMD Ryzen Threadripper 3970X
GPU	GeForce RTX 2080Ti
内存	250G
Python	3.6.12
Pytorch	1.7.1

运行深度学习模型的参数如表6所示：

表6　深度学习模型参数

参数	参数值
epoch	100
max_seq_length	100
learning_rate	3e－5
batch_size	80
dropout	0.5

4.3　基准模型对比

为了验证本文提出的模型的有效性，我们加入了如下对照实验：

(1)BiLSTM-CRF[40]：用于解决序列标注任务的经典模型，最初被用于命名实体识别任务，由双向长短时记忆网络和条件随机场所组成，前者用来捕捉序列的双向语义依赖关系，后者用来控制相邻标签依赖关系。

(2)BERT-BiLSTM-CRF：利用预训练模型BERT提取语义信息，加入双向长短期记忆网络进行进一步的语义信息建模，最后加上CRF控制标签输出。

(3)IDCNN-CRF[41－42]：由F. Yu等提出的膨胀卷积架构，相比于普通卷积，可以在卷积尺寸不变的情况下扩大卷积操作的感受野，后来由E. Strubell等引入到自然语言处理领域，将4个相同的膨胀卷积块叠在一起，每个块都包含膨胀宽度分别为1、1、2的三层架构。

(4)BERT-IDCNN-CRF[43]：利用预训练模型BERT获取语义表示，膨胀卷积用于提取特征，最后加上CRF层控制输出。本文提出的BERT-MultiCNN模型与各基准模型效果对比如表7所示：

表7　基准模型对比

Model	P/%	R/%	F_1/%
BiLSTM-CRF	76.23	50.35	60.64
BERT-BiLSTM-CRF	82.35	77.05	79.61
IDCNN-CRF	70.43	55.61	62.15
BERT-IDCNN-CRF	79.32	75.85	77.55
BERT-MultiCNN	83.98	80.49	82.20

从表7对比实验的结果可以看出：①在RNN和CNN这两种模式的架构下，BERT均可以在语义信息提取当中起到至关重要的作用，预训练模型采用大规模语料学习的方式对文本上下文信息进行有效建模，可以极大地提升原始基础模型的效果，而基础模型的作用，则是对BERT提取的信息进行进一步扩展与整合。在事件抽取这一任务中，BERT对不同基础结构的提升程度对比是：BiLSTM(＋18.97%)＞IDCNN(＋15.40%)，说明传统的双向长短期记忆网络在加入预训练模型后，仍然具有较大的提升空间；②BiLSTM在对句子序列建模方面仍然有着天然优势，优于膨胀卷积IDCNN的特征提取效果，这说明，IDCNN虽然可以通过扩大感受野的方式增加局部感知能力，然而即使扩大后其只能限定在固定宽度的范围内，对于事件要素之间的特定关联却无法正确表达；③总的来看，我们提出的模型BERT-MultiCNN与上述经典模型相比，准确率、召回率以及F1值均有了不同程度的提升，相比当前最优模型BERT-BiLSTM-CRF的F1值提升了2.59%，说明这一架构可以高效解决事件抽取中事件类型识别和事件要素识别这两个关键任务，在事件联合抽取任务上有较为广阔的应用空间。

4.4　消融实验

为了分析我们所提出模型的每个模块在整个事件抽取任务中的效果贡献，本文加入了如下消融实验，实验结果如表8所示：

表 8 消融实验

Model	P/%	R/%	F_1/%
Our_Model	83.98	80.49	82.20
-MultiCNN	81.24	79.43	80.32
-CRF	82.41	78.62	80.47
-Loss3	84.40	76.34	80.17

从以上实验结果可以看出：

（1）加入本文提出的多层 CNN 结构后，模型 F1 值提升了 1.88%，充分说明了后续 CNN 架构对编码层的语义捕捉进行了增强，可以有效捕捉到句子局部窗口和高层语义空间的信息，对于深层次语义建模具有促进作用。

（2）模型在加入 CRF 控制标签之间的转移后，准确率（+1.57%）和召回率（+1.87%）均有一定程度的提升，说明 CRF 在对事件要素标签进行建模方面具有一定的作用，然而受限于预训练模型的强大效果，其提升能力有限。

（3）去除了有效连接两个任务之间的对比损失之后，召回率急速下降 4.15%，然而准确率相差不大，说明加入的对比损失的确可以充分利用第一阶段事件句分类的信息，在减少错误传播的同时可以帮助模型识别出更多的事件实体，这对于一些需要高召回的场景具有很好的优化作用。

4.5 抽取结果讨论

经实验验证，本文提出的基于多层卷积神经网络的金融事件联合抽取方法，可以在没有其他辅助信息的条件下对原始新闻文本进行快速且准确的抽取，得到标准的结构化事件信息，抽取实际效果如表 9 所示：

表 9 抽取结果展示

输入文本	事件分类	事件元素抽取结果
2020 年 4 月 29 日上午 10 点，旅游学院与开封清明上河园股份有限公司战略合作签约仪式活动在清园会议室成功举办	1/战略合作	Trigger（触发词）：战略合作 Company（合作公司）：旅游学院 Company（合作公司）：开封清明上河园股份有限公司 Time（时间）：2020 年 4 月 29 日上午 10 点 Location（地点）：清园会议室
鸭河口水库自 9 月 25 日 4 时起加大下泄流量，泄洪流量将达白河最大行洪能力，南阳市中心城区低洼处有较大可能形成内涝	0/无关注事件	None
2020 年，杉杉股份布局 LCD 偏光片产业，募资收购 LG 化学偏光片资产，成为 LCD 偏光片的行业龙头	2/收购重组	Trigger（触发词）：募资收购 Company-1（收购主体）：杉杉股份 Company-2（收购客体）：LG 化学 Time（时间）：2020 年
2020 年底享道出行获得阿里巴巴和宁德时代 3 亿元 A 轮战略融资，此次融资主要用于发展企业级业务	3/投融资	Trigger（触发词）：融资 Company-1（投融资主体）：阿里巴巴 Company-1（投融资主体）：宁德时代 Company-2（投融资客体）：享道出行 Time（时间）：2020 年底 Round（轮次）：A 轮 Amount（金额）：3 亿元

同时，为了进一步说明本文模型对于事件抽取两阶段错误传播的改进作用，我们加入了进一步对照实验，对比不同模型在事件分类和元素抽取两个子任务中的实际效果，抽取结果见表 10。

根据表 10 结果可以看出，以往模型在解决金融事件抽取两阶段子任务时，事件元素抽取的结果强依赖于第一步骤中事件分类的结果，当事件分类错误（如 BiLSTM-CRF、IDCNN-CRF）时，后序元素抽取便毫无意义，而本文提出的金融事件联合抽取方法 BERT-MultiCNN，可以在保证抽取准确率（对比 BERT-BiLSTM-CRF、BERT-IDCNN-CRF）的同时，融合第二阶段中元素抽取的结果信息对事件分类环节进行错误纠正，以达到联合学习的目的，实现整体更优效果。

5 结语

本文聚焦于金融事件抽取中事件分类和事件要素抽取这两大子任务，针对已有研究中对卷积神经网络网络利用不充分、较少利用预训练模型提取语义信息、两个子任务之间错误传播这几大核心问题，提出了融合预训练模型与多层卷积神经网络的金融事件联合抽取框架 BERT-MultiCNN，先通过预训练编码模型 BERT 进行初级语义特征提取，将得到的句向量经处理后直接用于事件句分类，得到的字向量编码经多层卷积神经网络处理后，得到隐藏层字向量编码，并用于事件要素抽取任务，除了两个子任务进行损失的叠加训练之外，还建立了句向量表示上的联系，以进一步强化子任

表 10　不同模型抽取结果对比

输入文本	模型结构	事件分类	事件元素抽取结果
2020 年底享道出行获得阿里巴巴和宁德时代 3 亿元 A 轮战略融资，此次融资主要用于发展企业级业务	BiLSTM-CRF	1/战略合作	Trigger(触发词):战略 Company(合作公司):阿里巴巴 Company(合作公司):宁德时代 Company(合作公司):享道出行 Time(时间):2020 年底
	BERT-BiLSTM-CRF	3/投融资	Trigger(触发词):融资 Company-1(投融资主体):阿里巴巴 Company-2(投融资客体):享道出行 Time(时间):2020 年底
	IDCNN-CRF	0/无关注事件	None
	BERT-IDCNN-CRF	3/投融资	Trigger(触发词):融资 Company-2(投融资客体):享道出行 Time(时间):2020 年底 Round(轮次):A 轮 Amount(金额):3 亿元
	BERT-MultiCNN	3/投融资	Trigger(触发词):融资 Company-1(投融资主体):阿里巴巴 Company-1(投融资主体):宁德时代 Company-2(投融资客体):享道出行 Time(时间):2020 年底 Round(轮次):A 轮 Amount(金额):3 亿元

务之间的合理关联。本文设计的一系列相关对照实验可以说明，文中提出的联合抽取方法在金融事件抽取任务上达到了最优性能，可以应用于金融领域文本中的事件捕捉与要素抽取任务。同时，该方法也同样适用于其他领域的事件抽取。

本文提出的联合抽取模型当前主要用于单事件抽取。在更复杂的领域应用场景中，还有对语句中出现多个相关事件进行复合抽取的需求，我们将在后续工作中，探讨引入文档篇章信息对模型进行一步优化，以满足复合事件抽取需求。

参考文献：

[1] AHN D. The stages of event extraction[C]//Proceedings of the workshop and annotating and reasoning about time and events. USA: Association for Computational Linguistics, 2006: 1－8.

[2] CHEN C, NG V. Joint modeling for chinese event extraction with rich linguistic features[C]//Proceedings of COLING 2012. Mumbai: The COLING 2012 Organizing Committee, 2012: 529－544.

[3] DODDINGTON G, MITCHELL A, PRZYBOCKI M A, et al. The automatic content extraction (ace) program-tasks, data, and evaluation[J]. Proc Lrec, 2004, 2(1): 837－840.

[4] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the North American Chapter of the Association for Computational Linguistics: human language technologies. Minneapolis: The NAACL-HLT Press, 2019: 4171－4186.

[5] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 conference on empirical methods in natural language processing. Doha: Association for Computational Linguistics,2014: 1746－1751.

[6] RILOFF E. An empirical study of automated dictionary construction for information extraction in three domains[J]. Artificial intelligence, 1996,85(1/2): 101－134.

[7] RILOFF E. Automatically generating extraction patterns from untagged text[C]//Proceedings of the national conference on artificial intelligence. Oregon: Association for the Advancement of Artificial Intelligence,1996: 1044－1049.

[8] FELDMAN R, ROSENFELD B, BAR-HAIM R, et al. The stock sonar-sentiment analysis of stocks based on a hybrid approach[EB/OL]. [2021－11－10]. https://www.researchgate.net/publication/221016483_The_Stock_Sonar_-_Sentiment_Analysis_of_Stocks_Based_on_a_Hybrid_Approach.

[9] 罗明，黄海量. 基于词汇－语义模式的金融事件信息抽取方法[J]. 计算机应用, 2018,38(01): 84－90.

[10] 李响，杨小琳，魏勇，等. 基于支持向量机的新闻事件类型识别[J]. 地理信息世界, 2019,26(02): 73－78.

[11] HOU L, LI P, ZHU Q, et al. Event argument extraction based on CRF[C]// Proceedings of the 13th Chinese conference on Chinese lexical semantics. Berlin: Springer, 2012: 32－39.

[12] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 1: long papers). Beijing: Association for Computational Linguistics,2015: 167－176.

[13] ZENG Y, YANG H, FENG Y, et al. A convolution biLSTM neural network model for chinese event extraction[C]//NLPCC-IC-

CPOL 2016. Kunming: Springer, 2016: 275 - 287.
[14] 陈斌, 周勇, 刘兵. 基于卷积双向长短期记忆网络的事件触发词抽取[J]. 计算机工程, 2019, 45(01): 153 - 158.
[15] 吴文涛, 李培峰, 朱巧明. 基于混合神经网络的实体和事件联合抽取方法[J]. 中文信息学报, 2019, 33(08): 77 - 83.
[16] NGUYEN T H, CHO K, GRISHMAN R. Joint event extraction via recurrent neural networks[C]//Proceedings of the 2016 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies. San Diego: Association for Computational Linguistics, 2016: 300 - 309.
[17] 陈斌. 基于长短期记忆网络的事件抽取研究与应用[D]. 徐州: 中国矿业大学, 2019.
[18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//NIPS'17. New York: Curran Associates, 2017: 6000 - 6010.
[19] ZHENG S, CAO W, XU W, et al. Doc2EDAG: An end-to-end document-level framework for chinese financial event extraction [C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing. Hong Kong: Association for Computational Linguistics, 2019: 337 - 346.
[20] YANG S, FENG D, QIAO L, et al. Exploring pre-trained language models for event extraction and generation[C]//Proceedings of the 57th annual meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019: 5284 - 5294.
[21] ZHAO L, LI L, ZHENG X. A BERT based sentiment analysis and key entity detection approach for online financial texts[C]//Proceedings of the 2021 IEEE 24th international conference on computer supported cooperative work in design, 2021: 1233 - 1238.
[22] DU X, CARDIE C. Event extraction by answering (almost) natural questions[C]//Proceedings of the 2020 conference on empirical methods in natural language processing. Online: Association for Computational Linguistics, 2020: 671 - 683.
[23] NGUYEN T H, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection[C]//AAAI. Louisiana: Association for the Advancement of Artificial Intelligence, 2018: 5900 - 5907.
[24] CUI S, YU B, LIU T, et al. Event detection with relation-aware graph convolutional neural networks. [J]. CoRR, 2020, abs/2002.10757.
[25] YANG H, CHEN Y, LIU K, et al. DCFEE: A document-level chinese financial event extraction system based on automatically labeled training data[C]//Proceedings of ACL 2018, System Demonstrations. Melbourne: Association for Computational Linguistics, 2018: 50 - 55.
[26] EIN-DOR L, GERA A, TOLEDO-RONEN O, et al. Financial event extraction using wikipedia-based weak supervision[J]. ArXiv, 2019, abs/1911.10783.
[27] ZHOU Z, MA L, LIU H. Trade the event: corporate events detection for news-based event-driven trading[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Online: Association for Computational Linguistics, 2021: 2114 - 2124.
[28] RÖNNQVIST S, SARLIN P. Bank distress in the news: describing events through deep learning[J]. Neurocomputing, 2017, 264: 57 - 70.
[29] CARTA S, CONSOLI S, PIRAS L, et al. Event detection in finance using hierarchical clustering algorithms on news and tweets [J]. PeerJ computer science, 2021, 7: 438.
[30] CORRO L D, HOFFART J. Unsupervised extraction of market moving events with neural attention[J]. ArXiv, 2020, abs/2001.09466.
[31] ZHANG Y, WALLACE B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[C]//Proceedings of the eighth international joint conference on natural language processing (volume 1: long papers). Taipei: Asian Federation of Natural Language Processing, 2017: 253 - 263.
[32] XU H, LIU B, SHU L, et al. Double embeddings and CNN-based sequence labeling for aspect extraction [C]//Proceedings of the 56th annual meeting of the Association for Computational Linguistics (volume 2: short papers). Melbourne: Association for Computational Linguistics, 2018: 592 - - 598.
[33] MA X, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[C]//Proceedings of the 54th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Berlin: Association for Computational Linguistics, 2016: 1064 - 1074.
[34] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing. Doha: Association for Computational Linguistics, 2014: 1532 - 1543.
[35] NGUYEN T H, GRISHMAN R. Event detection and domain adaptation with convolutional neural networks[C]//Annual meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Conference. Beijing: Association for Computational Linguistics, 2015: 365 - 371.
[36] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. ArXiv, 2013, abs/1301.3781.
[37] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541 - 551.
[38] BOUVRIE J. Notes on convolutional neural networks[EB/OL]. [2021 - 11 - 10]. http://cogprints.org/5869/1/cnn_tutorial.pdf.

[39] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]//Proceedings of the eighteenth international conference on machine learning. San Francisco: Morgan Kaufmann Publishers, 2001: 282－289.
[40] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]//Proceedings of the 2016 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies. San Diego: Association for Computational Linguistics, 2016: 260－270.
[41] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. CoRR, 2016, abs/1511.07122.
[42] STRUBELL E, VERGA P, BELANGER D, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. Copenhagen: Association for Computational Linguistics, 2017: 2670－2680.
[43] 李妮，关焕梅，杨飘，等. 基于BERT-IDCNN-CRF的中文命名实体识别方法[J]. 山东大学学报(理学版)，2020,55(01): 102－109.

作者贡献说明：

李旭晖：提出研究思路，对研究方法提供指导，修改论文；
程威：采集、清洗和分析数据，实现研究方案，编写代码，撰写论文，修改论文；
唐小雅：实现研究方案，修改论文；
于滔：数据预处理，对研究过程提供指导，修改论文；
陈壮：对研究过程提供指导，提出修改意见；
钱铁云：对研究过程提供指导，修改论文。

A Joint Extraction Method of Financial Events Based on Multi-Layer Convolutional Neural Networks

Li Xuhui[1,2] Cheng Wei[1] Tang Xiaoya[3] Yu Tao[1] Chen Zhuang[3] Qian Tieyun[3]
[1] School of Information Management, Wuhan University, Wuhan 430072
[2] Big Data Institute, Wuhan University, Wuhan 430072
[3] School of Computer Science, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] In order to further improve the effect of event extraction in the financial field, the correlation between the two subtasks of event extraction needs to be enhanced. [**Method/process**] This paper carried out related research about event extraction on Chinese financial texts, and proposed a joint extraction method of financial events that integrated the pre-training model and a multi-layer convolutional neural network. First, the pre-training model BERT captured the comprehensive semantic information of the sentence sequence, then accessed the multi-layer convolutional architecture designed in this paper——MultiCNN, hierarchically extracted local window and high-dimensional spatial semantic information, realized the two tasks of event recognition and element extraction at the same time, and then introduced contrast loss to further strengthen the association between the two tasks. [**Result/conclusion**] F1 has reached 82.20% on the Chinese financial event data set, which has a certain improvement over the benchmark extraction models.

Keywords: Chinese event extraction convolution neural network pretraining model joint learning